

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2002, 13:35:49 ; Search time 19.47 seconds  
(without alignments)  
2527.415 Million cell updates/sec

Title: US-09-405-504A-25  
Perfect score: 3372  
Sequence: 1 MRAPGAGAAVSVSLALLWLL.....HYLPLNEAVVTRICSGAFAL 646

Scoring table: BLOSUM62  
Gap 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues  
Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_68: \*  
1: pir1: \*  
2: pir2: \*  
3: pir3: \*  
4: pir4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	3054	90.6	646	1 A55093	fatty acid transpo
2	1267.5	37.6	650	2 T21498	hypothetical prote
3	1257.5	37.3	655	2 T15879	hypothetical prote
4	988.5	29.3	689	1 JW0107	very-long-chain ac
5	899.5	26.7	597	2 D70609	probable fadD6 pro
6	872	25.9	608	2 H83284	probable very-long
7	747.5	22.2	643	2 T43052	fatty acid transpo
8	699	20.7	623	2 S45899	probable membrane
9	406.5	12.1	522	2 H85484	probable crotonob
10	390.5	11.6	522	2 S40558	probable carnitine
11	377	11.0	502	2 D70806	probable fadD17 pr
12	372	11.0	532	2 G70986	probable coA ligas
13	361	10.7	2723	2 T03221	probable polyketid
14	346.5	10.3	503	2 E70853	probable acid--CoA
15	325	9.6	444	2 T50931	hypothetical prote
16	323.5	9.6	549	2 E69893	probable acid--CoA
17	320	9.5	593	2 E69378	probable acid--CoA
18	319	9.5	549	2 T44812	probable fatty-aci
19	316	9.4	513	2 A69831	probable acid--CoA
20	315	9.3	999	2 B70501	probable fatty-aci
21	307	9.1	486	2 H69656	hypothetical prote
22	305	9.0	560	2 A70628	O-succinylbenzoate
23	301.5	8.9	2560	1 T40457	probable acid--CoA
24	300.5	8.9	599	2 H72454	peptide synthetase
25	299	8.9	3588	2 T40485	probable fatty-aci
26	298.5	8.9	544	1 S01667	surfactin syntheta
27	297.5	8.8	544	2 S15695	4-coumarate--CoA 1
28	294	8.7	514	2 T46131	4-coumarate--CoA 1
29	291.5	8.6	8563	2 T30226	polyketide synthas

30	289.5	8.6	557	2	T07909
31	289.5	8.6	4735	2	T17463
32	288.5	8.6	569	2	C69471
33	287.5	8.5	3587	2	T40486
34	286.5	8.5	535	2	T08074
35	286.5	8.5	552	2	E69438
36	286.5	8.5	566	2	D85778
37	284.5	8.4	493	2	G84263
38	282.5	8.4	548	2	T07908
39	281.5	8.3	546	2	C83791
40	281.5	8.3	561	2	D96674
41	281	8.3	563	1	JU0311
42	280.5	8.3	566	2	E64928
43	278.5	8.3	545	1	B39827
44	277.5	8.2	545	1	A39827
45	277	8.2	551	2	H69371

ALIGNMENTS

RESULT 1

A55093

fatty acid transport protein precursor - mouse

C:Species: Mus musculus (house mouse)

C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 08-Sep-2000

C:Accession: A55093; I49132

R:Schaffer, J.E.; Lodish, H.F.

Cell 79, 427-436, 1994

A:Title: Expression cloning and characterization of a novel adipocyte long chain  
A:Reference number: A55093; MUID:95042740

A:Accession: A55093

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-646 <SCH>

A:Cross-references: GB:U15976; NID:g563828; PIDN:AAC71060.1; PID:g563829

C:Genetics:

A:Gene: FATP

C:Superfamily: Mycobacterium tuberculosis probable fadD6 protein; acetate--CoA lig  
F:124-604/Domain: acetate--CoA ligase homology <ACLU>

Query Match 90.6%; Score 3054; DB 1; Length 646;

Best Local Similarity 89.5%; Pred. No. 1.9e-237;

Matches 578; Conservative 30; Mismatches 38; Indels 0; Gaps 0;

QY 1 MRAPGAGAAVSVSLALLWLLGLPWTWSAAAALGVYVGGGWRFLRVCKTARRDLFGLSV 60

Db 1 MRAPGAGTASVASLALLWLLGLPWTWSAAAALGVYVGGGWRFLRVCKTARRDLFGLSV 60

QY 61 LIRVRLRRHORAGHTIPRIFQAVVQRPRLALVDAGTGCWTFQAQLDAYSNVANL 120

Db 61 LIRVRLRRRRAGDTPICIFQAVARRQPERLALVDAGTGCWTFQAQLDAYSNVANL 120

QY 121 RQLGFAPGDVVAIFLEGPRFVGLWGLAKAGMEALLNVRERPLAFLCTGSGAKALI 180

Db 121 RQLGFAPGDVVAIFLEGPRFVGLWGLAKAGVAAALLNVRERPLAFLCTGSGAKALI 180

QY 181 FGGEMAAVAEVSGLKSLIKFCSDGLGPEGLPTDTHLLDPLLEKASTAPLAQIPSKGM 240

Db 181 YGGEMAAVAEVSGLKSLIKFCSDGLGPEGLPTDTHLLDPLLEKASTAPLAQIPSKGM 240

QY 241 DDLRFYIYTSCTTGLPKAAIVHSRYRMAAFHGHYRMAADVLVDCPLPLVHSAAGNI 300

Db 241 DDLRFYIYTSCTTGLPKAAIVHSRYRMAAFHGHYRMAADVLVDCPLPLVHSAAGNI 300

QY 301 VGQCLYGTIVLVRKKFSRFRWDDCIKYNCTVQVQVIGETICRYLLKQPVREARRHRVL 360

Db 301 VGQCLYGTIVLVRKKFSRFRWDDCIKYNCTVQVQVIGETICRYLLKQPVREARRHRVL 360

QY 361 AVNGLRPAIWEETFRFGRVQIGEFYATECNCSIANMDGVGSGGFSRILPHVYPIR 420

Db 361 AVNGLRPAIWEETFRFGRVQIGEFYATECNCSIANMDGVGSGGFSRILPHVYPIR 420



Db 433 CELERDKNGLCVPCVGETGEMVGIKEKDILLKFEYVSEGDFTAKKIYRDVFKHGDKVF 492  
QY 489 LSGDVLWDELGYMYFDRSDGTPFRGENVSTTEVEGVLSRLLGQDTPDAVYGVAVPGVE 548  
Db 493 ASGDILLHWDLLGLYLFVDRCDGTPFRMGENVSTTEVEGILQPMVDVEDATVYGVTVGKME 552  
QY 549 GKAGMAA-----VADPHSLDDNAYIQEQLQKVLAPARPIFLRLLLPOVDTTGT 596  
Db 553 GRAGMAGIVKDGTDVKEFTADITSRLTEN-----LASVAIPVFIRLCKEVDRTGT 603  
QY 597 FKIQKTRLQREGFD-PRQTSRLRFLFDLQKQHYLPLNEAVYTRICSGAF 644  
Db 604 FKLKKTDLQOGYDLVACKGDPIYWSAAEKSYPKPLTDKMQQDIDTGVY 652

RESULT 4  
JW0107  
very-long-chain acyl-CoA synthetase related protein - mouse  
N:Alternate names: VLACSR  
C:Species: Mus musculus (house mouse)  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 08-Sep-2000  
C:Accession: JW0107  
R:Berger, J.; Truppe, C.; Neumann, H.; Forss-Petter, S.  
Biochem. Biophys. Res. Commun. 247, 255-260, 1998  
A:Title: A novel relative of the very-long-chain acyl-CoA synthetase and fatty acid tran-  
sacylase from the liver of the house mouse.  
A:Reference number: JW0107; MUID:98308102  
A:Accession: JW0107  
A:Molecule type: mRNA  
A:Residues: 1-689 <BER>  
A:Cross-references: GB:AJ223959  
A:Experimental source: Liver  
C:Comment: This protein likely functions as a plasma membrane transporter of long chain  
acids.  
C:Superfamily: Mycobacterium tuberculosis probable fadD6 protein; acetate--CoA ligase ho-  
mology <ACL>  
F:169-647/Domain: acetate--CoA ligase homology <ACL>

Query Match 29.3%; Score 988.5; DB 1; Length 689;  
Best Local Similarity 37.5%; Pred. No. 2.6e-71;  
Matches 260; Conservative 105; Mismatches 243; Indels 85; Gaps 20;

QY 10 SVVSLALLW-----LGLPW--TWS-----AAALGVYV-----GSGGWRP 43  
Db 26 AAMALALRWFLGDPTCLVLLGLALGRPWISSMHWLSLVGAALTFLPLQPPPLGRW 85  
QY 44 L-RIVCKTARRDLFGLSVLRVRLRHRORAGHTIPRIFQAVVQRQ-----PERLALYDA 98  
Db 86 LHKDVAFTKMLFYGL-----KFRRLNKH-----PETFDALERQALAWDRVALVCT 135  
QY 99 GT-GEQWTFQAIDAYS-----NAVANLFRQLGAPGDVVVAIFL--EGRPEFVG 143  
Db 136 GSEGSITSQDARSQAANWLKAKLKDAVIONTR-----DAAAILVPSKTSIALS 188  
QY 144 LWTGLKAGMEALLNVNLRREPLAFCLGTSGAKALIFGEMVAVAEVSQH-LGKSLIK 202  
Db 189 VFLGLAKLCPVAVINPHSRGMPHLHSVRSSGASVLVDDQENLEEVLPKLLAENIHC 248  
QY 203 FCGSDGLPGGILPDTHLLDPLKKEASTAPL-----AQIPSKGMDRLFYITSGTGLPK 257  
Db 249 FYLGHSSP---TPGVGALGASLDAAPSDVPASLRATIKW---SPAIFITSGTGLPK 302  
QY 258 AATVHSRYRMA---AFGHAYRMAQADVLDCLPLVHSAGNIIGVQCCLYIGTVVLR 314  
Db 303 PALLSHERVIOVSNLSF---CGCRADDVYVDVPLVHTIGLVGLGCLQVGATCVLA 358  
QY 315 KKFASRFRWDCDCKYVCTVVOYIGETCRYLLKQPVREARRHRVRLAVNGNLRPAIWEF 374  
Db 359 PFSASRFAEACRQHGVTILVGEILRYLCNVPEQEDKHITVRLAWNGNLRANWKNF 418  
QY 375 TERFGVRIQTEFGATECNCSIANMDGKVGCGFNRSRILPHVYIRLVKVNEDTMEILLRD 434  
Db 419 QORFGPIRIWEFVSGTEGNVGLMNVVHCGAVGRTSCILRLMTPPELVQFDIETAEPLRD 478

QY 435 AOGCLIPCOAGBGLLVGQINQODPLRRFDGYVSESATSKKTAHVSFKSGDSAYLSGDVL 494  
Db 479 KQGFCLPVPBPGKGLLLTKVRKNQPLGYRG--SQAESNRKLVANVRRVGDLYFNTGDVL 536  
QY 495 VMDLGYMYFRDSGDTFRWGENVSTTEVEGVLSRLLGQDTPDAVYGVAVPGVEGKAGMA 554  
Db 537 TLDQEGFFYFQDRLGDTFRWGENVSTTEVEGVLSRLLGQDTPDAVYGVAVPGVEGKAGMA 596  
QY 555 AV-ADPHSLDDNAYIQEQLQKVLAPARPIFLRLLLPOVDTTGTFRKIQKTRLQREGFDPRQ 613  
Db 597 AVKLAPGKTFDQKLYQHVRSWLPAYATPHFIRIODSLTEITNTYKLVRSRLVREGFDVGI 656  
QY 614 TSDRLFFDLKQHYLPLNEAVYTRICSGAFAL 646  
Db 657 IADPLYILDNKAQTFERSLMPDVYQAVCEGTWKL 689

RESULT 5  
D70609  
probable fadD6 protein - Mycobacterium tuberculosis (strain H37Rv)  
C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 18-Aug-2000  
C:Accession: D70609  
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, N.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holt,  
Nature 393, 537-544, 1998  
A:Authors: Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.;  
Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete ge-  
nome sequence.  
A:Reference number: A70500; MUID:98295987  
A:Accession: D70609  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-597 <COL>  
A:Cross-references: GB:293777; GB:AL123456; NID:g3261726; PIDN:CAB07829.1; PID:g19  
C:Superfamily: Mycobacterium tuberculosis probable fadD6 protein; acetate--CoA lig-  
ase homology <ACL>  
F:93-555/Domain: acetate--CoA ligase homology <ACL>

Query Match 26.7%; Score 899.5; DB 2; Length 597;  
Best Local Similarity 36.6%; Pred. No. 3.1e-64;  
Matches 222; Conservative 94; Mismatches 251; Indels 39; Gaps 12;

QY 35 YVSGGWRFLRIVCKTAR--RDLFGLSVLR---VRLRHRORAGHTIPRIFQAVVQRQ 89  
Db 4 YYG-GAHTTVRLDLATMPRVLADTPVIVRGAMTGLLARNPSKA--SIGTVFDRAARY 60  
QY 90 PERLALVDAGTGCWCTFAQLDAYSNVANLFRQLGAPGDVVVAIFLEGRPEFVGLMLGLA 149  
Db 61 GDRVFLFEG--DQGLTTRDANATANYAAVLAARGVPGDVGVGIMLRNPSSTVLAMLATV 118  
QY 150 KAGMEALLNVNLRREPLAFCLGTSGAKALIFGEMVAVAEVSQHGLKSLIKFCSCDGLG 209  
Db 119 KCGAIGMLNVHQREGLVLAHSLGLLDKAVLIAESDILVSAVAECCASGR-----167  
QY 210 PEGILPDTHLLDPLKKEASTAPL---AQIPSKGMDRLFYITSGTGLPKAAIVHSRY 266  
Db 168 ---VAGDVLTAVEDVERFATTAPATNPASASAVQAKDTAFYITSGTGTFFPKASVMTTHRW 224  
QY 267 YR-MAAFGHAYRMAQADVLDCLPLVHSAGNIIGVQCCLYIGTVVLRKFKFSASRPWDD 325  
Db 225 LRLAVFGGMLRUGSDTLVSCPLVHNNALTVAVSSVINSNGATLALGKFSASRPWDE 284  
QY 326 CIKYNCTVVOYIGETCRYLLKQPVREARRHRVRLAVNGNLRPAIWEETERGCVROIGE 385  
Db 285 VIANRATAFVYIGETCRYLLNQPAPKPTDRAHOVRVVCNGLRPEIWEDEFTTRFGVARVCE 344  
QY 386 FYGATECNCSIANMDGKVGCGFNRSRILPHVYIRLVKVNEDTMEILLRDAGGLCIPQAG 445  
Db 345 FYAASEGNSAFINFPRTAGVSP-----MPLAFVEYDLDITGDPDLRDSGRVRRVPDG 398



Db 542 RVRSINCHGRRRDRVDIRFAALAKHARDRLPGVAVPLFLRVTPALEYTGTLKIQKGLKQ 601  
QY 607 EGPDPRTS--DRLFFDLKOGHYPLNE 633  
Db 602 EGIDPDKISGEDKLYLWLPGPSDIYLPFGK 630

## RESULT 8

S45899  
probable membrane protein YBR041w - yeast (Saccharomyces cerevisiae)  
N/Alternate names: hypothetical protein YBR0411  
C/Species: Saccharomyces cerevisiae  
C/Date: 26-Aug-1994 #sequence\_revision 09-Sep-1994 #text\_change 02-Sep-2000  
C/Accession: S45899  
R/Andre, B.; Cziepluch, C.; Hein, C.; Jauniaux, J.C.; Urrestarazu, A.; Visser, S.  
submitted to the Protein Sequence Database, August 1994  
A/Reference number: S45899  
A/Accession: S45899  
A/Molecule type: DNA  
A/Residues: 1-623 <AND>  
A/Cross-references: EMBL:235910; NID:g536263; PIDN:CAA84983.1; PID:g536264; MIPS:YBR041w  
A/Experimental source: strain S288C  
C/Genetics:  
A/Gene: SGD:FAT1  
A/Cross-references: SGD:S0000245; MIPS:YBR041w  
A/Map position: 2R  
C/Superfamily: Mycobacterium tuberculosis probable fadD6 protein; acetate--CoA ligase h  
C/Keywords: transmembrane protein  
F;1-27/Domain: transmembrane #status predicted <TM1>  
F;55-71/Domain: transmembrane #status predicted <TM2>  
F;133-623/Domain: acetate--CoA ligase homology <ACL>  
F;149-167/Domain: transmembrane #status predicted <TM3>  
F;304-322/Domain: transmembrane #status predicted <TM4>

Query Match 20.7%; Score 699; DB 2; Length 623;  
Best Local Similarity 33.5%; Pred. No. 4.4e-48;  
Matches 190; Conservative 87; Mismatches 234; Indels 56; Gaps 19;

QY 66 LEURHORAGHTIPRFQAVVOQPERLA-----LVDAAGTCEWTFQAOLDAYSNAVAML 119  
Db 69 IDVRHRFQNW---YLFQVQOQGDHLAISYTRPMAEKGFEQLETFYIETV-NIVLRL 124  
QY 120 FROLGF----APGDVVAIFLEGRPEFVGLWGLAKAGMEAAALNNLRREPLAFCLGTS 175  
Db 125 SHLLHEDYNVQAGDYVAIDCTNKPLFVFLWLSLWNGAIPALNVTGCTPLVHSLKISN 184  
QY 176 AKALIFGGEVAAVAVSVSHLGLKSLKFCGDLGPE---GILPDTHLLDPLLKEASTAPL 232  
Db 185 ITQVTFIDPASNPIRESEEEIKNAL-----PDVKLNYLEEQDLMHELLNSQSPFL 235  
QY 233 AQIPSK---GMD--RLFYIYTSGLPKAAIVVHSRYRMAAFG---HHAYBQAAD 283  
Db 236 QQDNVTRPLGLTDFKPSMLIYTSGLTGLPKSAIM---SWRKSSVGCQVFGVHLWMTNES 291  
QY 284 VLYDCLPLVHSAAGNIGVQCCLYGLTVVLRKKFSASRPWDCIKYNTVVQYIGEIRY 343  
Db 292 TVFTAMPPLFHSAAALIGACAILSHGCLALSHKFSASTFWKQVLTGATHIQYVGEVCY 351  
QY 344 LLKQPVREARRHRVLA VNGRLPAINEEFTEFREGVRIQIEFFYGATE---CNCSIAND 400  
Db 352 LLHTPLSKYEKMKVAVNGNGLRPIWODFRKRNIEVIEFGEYAATAPFATTTFOKGD 411  
QY 401 GKVGSGFNRSLPHWYPIR--LVKV--NEDTMELLRDAQGLICPCQAGEPG-LLVGGIN 455  
Db 412 FGIGACRNYGTIIQWFLSQOQLVRMDPNDSDV-YLRNKGCEVAVPGEFGEMLRIF 470  
QY 456 QODPLRDFGYVSES-ATSKTAHSVFSKGSAYLSGDLVMDLGYMYFRDRSGDTPRW 514  
Db 471 PKKPETSFGYLGNAKETSKVVRDVRFRGDAWYRCGLLKADEYGLWYFLDRMGDTPRW 530  
QY 515 RGENYSTTEVEGLV--SRLLGQTDVAVYGVAVPGVEKGAGNAVADPHSLLDPA-----I 568

Db 531 KSENVSTTEVEDQLTASNKEQYQAQVLVVGIKVPKEGRAGFAVIKLTNSLSDITAKTKLL 590  
QY 569 YOELOKVLAP-YARPIELRLLPQVDTT 594  
Db 591 NDLSRLNLPSYAMPFLVKFVDEIKMT 617

## RESULT 9

H85484  
probable crotonobetaine/carnitine-CoA ligase caic [imported] - Escherichia coli (s  
C/Species: Escherichia coli  
C/Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 31-Mar-2001  
C/Accession: H85484  
R/Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.;  
Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; A  
Nature 409, 529-533, 2001  
A/Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A/Reference number: A85480; MUID:21074935; PMID:11206551  
A/Accession: H85484  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-522 <STO>  
A/Cross-references: GB:AE005174; NID:gl2512724; PIDN:AAG54340.1; GSPDB:GN00145; UW  
A/Experimental source: strain O157:H7, substrain EDL933  
C/Genetics:  
A/Gene: caic  
C/Superfamily: 4-coumarate--CoA ligase; acetate--CoA ligase homology

Query Match 12.1%; Score 406.5; DB 2; Length 522;  
Best Local Similarity 25.9%; Pred. No. 1.1e-24;  
Matches 139; Conservative 85; Mismatches 249; Indels 63; Gaps 16;

QY 92 RLALV---DAGTGEWTFQAOLDAYSNAVANLFRQLGAPGDVVAIFLEGRPEFVGLWGL 148  
Db 28 KTAICESGGVNVRYSVLELNQNEINRTANFLYTLGTRKGDVLAHLNDCPEFIFCMFGL 87  
QY 149 AKAGMEAAALNNLRREPLAFCLGTSGAKALIFGGEVAAVAVSVSHLGLKSLKFCSGDL 208  
Db 88 AKIGAIMVPIINARLREESAWILQNSOACLIVTSAQFVPMYQQIQQEDATOLRHICLTDV 147  
QY 209 ---GPEGLPDPHLLDPLLEKASTAPLAQIPSKGMDRLFYIYTSGLTGLPKAAIVVHSR 265  
Db 148 ALPADDGVSSFTQ-----LKNQQPATLCYAPPLSTDDELTAEILFTSGTTSRPGVVIITH-- 200  
QY 266 YYRMAAFGHHA---YRMAADVLVDCLPLVHSAAGNIGVQCCLYGLTVVLRKKFSASRF 322  
Db 201 -YNLRFAGYYSAMQCARDDDDVLTVPAPAFHIDCQCATAAFAFSAGATFVLVEKYSARAF 259  
QY 323 WDDCIKYNCTVVQYIGICRYLLKQPVREARRHRV-LAVNGRLRPAINEEFTEFREGVR 381  
Db 260 WGVOKYRATITECIPMIRTLVMOVPPSANDRQHLREVFMFLNLSQERKDFCERFGR 319  
QY 382 QIGEFYGATECNCISAN-----MDGKVGSCGNSRIL-PHYVPILRVLVNEDTME 430  
Db 320 LLTS-YGMTEFIVGIDRPGCDKRWPSIGRAGFC-YDAEIRDDHNRPLPAGEIGE---- 373  
QY 431 LLRDAQGLICPCQAGEPCLLVQINQODPLRRFDGYVSESATSKTAHSVFSKGSAYLS 490  
Db 374 -----ICI---KGVPGKTIK-----EYFLNPKATAKVL-----EADGWLHT 407  
QY 491 GDVLVMDLGVMYFRDRSGDTFRWGENVSTTEVEGLVRLLLGQTDVAVYGVAVPGVEKG 550  
Db 408 GDTGYRDEEGFFIFIDRRCNMRKGGNSVCELENIATHPKIQDVIWVGIR-DSIRDE 466  
QY 551 AGMA-AVADPHSLLDPAIYQELQKVLAPYARPIELRLLPQVDTTGTGFTQKTRLQ 605  
Db 467 ARAVAVVNLNGETLSEEEFFCFCEQNAKFKVSYLEIRKDLPRNCSGKIIRNLK 522

## RESULT 10

S40558



Db 415 TEVAVYVPVDPVGDVQVMAALVLAPEKTKEDADKFAFLTRQPDGLGHKQWPSYVVSAGLP 474

QY 593 TTGTFKIQKTRLOREGF---DP 611

Db 475 RTMTFKIKKQLSAGVACADP 496

## RESULT 12

G70986

probable coA ligase - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 18-Aug-2000

C:Accession: G70986

R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.

; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:98295987

A:Accession: G70986

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-532 <COL>

A:Cross-references: GB:295890; GB:AL123456; NID:G3242245; PIDN:CAR09316.1; PID:G2131015

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: fadD1

C:Superfamily: 4-coumarate--CoA ligase; acetate--CoA ligase homology

F:48-499/Domain: acetate--CoA ligase homology <AC1>

Query Match 11.0%; Score 372; DB 2; Length 532;

Best Local Similarity 26.2%; Pred. No. 6.9e-22;

Matches 141; Conservative 82; Mismatches 221; Indels 94; Gaps 19;

QY 104 WTAQLDAYSNVANVLFRLQGFAPGDVVAIFLEGRPEFVGLMLG-----LAKAGMBA 155

Db 27 WTSQVLAESAARA-----AALTIADPQRTHIGSLGNTPEMLAQLAAAGLGG 76

QY 156 AL---LNVNLRREPLAFCLGTSGAKALIFGGMVAVAEVSGLHKLKFCSGDLGPBG 212

Db 77 YVLCGLTTRGDALADVRRADCQIVTDADHRALL-----DG 115

QY 213 I-LPDTHLDD---PLLKASTAPLAQIPSGMD--DRLEFYIYTSGLTPKAAIVVHSRY 266

Db 116 LDLAGARILDTSPRAELVAGDGFVPYREVDTMDPFMMIIFTSGLSGNPKAVPVSH--- 172

QY 267 YRMAAFGHA---YRQADVLYDCLPLYHSAGNIIGVQCILYGLTVVLRKKSASRF 322

Db 173 -LMATFAGSLTERFGLTEQDTCYVSMPLFHSNNAVAGWAPVAVSG-AAIAPATFSATGF 230

QY 323 WDCIKYNTVVOYIGEICRYLLKQPVREARRHVRVLAAGNGLRPAITEEFERFGRQ 382

Db 231 LDDVRYHATYMYKPLAYILATPERDDADNPLURVAFGNEANDKJEEFSRRFGV-Q 289

QY 383 IGFYGATCNCNSIAMDG-KYVSCG-----FNSRLPHVPIRLVKVNDTWELLR 433

Db 290 VEDFGSTENAVIVIREPPTPGSIGRGAGHVAVINGEIVTECAVAF----- 337

QY 434 DAQGLIPCAGEPGLLVQINQDPLRFDFGVSE-SATSKIAHSVFSKGSAYLSGD 492

Db 338 DAHAL--TNADE---AIGELVNTTGSFFTGYNDPEANERMRHGMVWSDGLAY---- 388

QY 493 VLVMDLGYMYFRDRSGDTFRWGENVSTTEGVLSRLLGOTDVAVYGVAVPVEGKAG 552

Db 389 ---RDEGWIYLAGRTADMRVVDGENTAAPIERILLRYKAINRVAVYAVPDEYVGQVM 445

QY 553 MAADVPHSLDPNAT--YQELQKVLAPYARIFLRLLPQVDTTGTTFKIQKTRLOREG 608

Db 446 AALVLRAGDTFDPDAFEALDAQPDLSKARPYIRIAADLPSTATHKVLKRLIDEG 503

## RESULT 13

T03221

probable polyketide synthase module 1 - Streptomyces hygroscopicus

C:Species: Streptomyces hygroscopicus

C:Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 17-Nov-2000

C:Accession: T03221

R:Ruan, X.; Stassi, D.; Lax, S.; Katz, L.

Gene 203, 1-9, 1997

A:Title: A second type-I PKS gene cluster isolated from Streptomyces hygroscopicus

A:Reference number: 214848; MUID:98085969

A:Accession: T03221

A:Status: preliminary; translated from GB/EMBL/DBD

A:Molecule type: DNA

A:Residues: 1-2723 <RUA>

A:Cross-references: EMBL:AF007101; NID:G2624946; PIDN:AAC38061.1; PID:G2624948

A:Experimental source: ATCC 29253

C:Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I homology; acetate--CoA

C:Keywords: carrier protein

F:54-504/Domain: acetate--CoA ligase homology <AC1>

F:881-949/Domain: acyl carrier protein homology <ACP1>

F:996-1397/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS>

F:1509-1794/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT>

F:2568-2639/Domain: acyl carrier protein homology <ACP2>

## Query Match

10.7%; Score 361; DB 2; Length 2723;

Best Local Similarity 27.5%; Pred. No. 5.3e-20;

Matches 156; Conservative 85; Mismatches 266; Indels 60; Gaps 20;

QY 64 VRLELRHQRAGHTIPRIFOAVVORQPERLALVDAGTGCECTFAQLDAYSNVANLFRQL 123

Db 2 LRRELIR-----PLPELLKAHARRGQTAFADSRRCV--SYAQLEARTRLAGHLAQL 53

QY 124 GFAPGDVVAIFLEGRPEFVGLWGLAKAGMEALLNVNLRREPLAFCLGTSGAKALIFGG 183

Db 54 RLLPGDRAAIYLGNCVETIESYLATARASATGVPINPHSSSTAYELAYLLDLSAARVIVTDH 113

QY 184 EMVAABAVSVGHGSLKIFCGSDLPBGILPDTHLLDPLKKEASTAPLAQIPSK---GM 240

Db 114 TRLQVRRLQAE-RPHLTVVVTG-D-EESGAVP-----FDAL---AETEP---QOPARDLGL 162

QY 241 DDRLFYIYTSGLTGLPKAAI--VVHSRYRYMAAFGHAYRMAQADVLYDCLPLYHSAGNII 299

Db 163 DDIAWMLYTSGLTGRKGVLTQRSCLSVSAASSAGVLGLSETDRVLWPVPLFHSIAHY 222

QY 300 GVGQCLYGLTVLVRKFSASRFDDCKYKNTCTVVQYIGETCRYLLKQPVREAE---RRH 356

Db 223 CVLSVTAVGATARITEAFDAEELLTLRAEEFTFLAGVPTMYHYLL-DAARDGDLSPNL 281

QY 357 RVLAVGNGLRPAIWEETFERFGRQIGEFYGAIE-CNCSTANMDGRV---GSCGFNSRI 412

Db 282 RVLCSAGAISTATIRAEFOETFGVPLL-DCVGSSTETGLITANMPNQTQEGSCG----- 335

QY 413 LPHYPIRLVKVNDTWELLRDAQGLCIPCQAGEPGLLVGINQDPLRRFDGTVVSESAT 472

Db 336 LP-VPLGTVTRIVDPETHESVETGEGEI--WVGPSLMVGVHNPQ-----ATEAAL 384

QY 473 SKKTAHSVFSKGSAYLSGDVLMDELGYMYFRDRSGDTFRWGENVSTTEGVLSRL 532

Db 385 PR-----GWYRTGDLGRDRDLGYHTITGRLSLIIIRSGENTHPTTEVEQVLLRVP 433

QY 533 GQTDVAVGVAVPVEGKAGMAAVADPSSLDDPNAIYQELQKVLAPYARIFLRLLPQVD 592

Db 434 GVTDAVVSNAHPTL-GEVPVAVLVPADGDPDPEFAACREHLAYFKVPEELYEIGSIP 492

QY 593 TTGTFTKIQKTRLOREGFDPQRTSDRLF 619

Db 493 RTGSGKIKRHALAQSAPRLRAVSSGSF 519

## RESULT 14

E70853

probable acid--CoA ligase (EC 6.2.1.-) fadD13 - Mycobacterium tuberculosis (strain

